

Лекція 9

Чисельні методи оптимізації алгоритмічно заданих критеріїв задач нелінійного програмування

Розділ 9.1. Загальна характеристика чисельних методів

Чисельні методи відіграють значну роль у прикладних дослідженнях. Це обумовлено рядом причин, серед яких головне місце займає різноманіття функцій $f(X)$ та $g_i(X), i = 1, m$, а також форм їх завдання. В окремих випадках навіть важко визначити, до якого класу відноситься та чи інші задача, а також чи існує для неї обґрунтований метод розв'язання.

Розроблено багато чисельних методів для задач як безумовної, так й умовної оптимізації. Природним є бажання вибрати для розв'язання конкретної задачі найкращий метод, який дозволяє отримати розв'язок із необхідною точністю з максимальним використанням потужності ЕОМ.

Якість чисельних методів характеризується багатьма факторами: швидкість збіжності, час виконання однієї ітерації, обсяг пам'яті ЕОМ, який є необхідним для реалізації методу, клас задач, що розв'язуються і т. ін. Розв'язувані задачі також дуже різноманітні: вони можуть мати велику чи малу вимірність, бути унімодальними чи багатоекстремальними тощо. Один і той же метод, що є ефективним для розв'язання задач одного типу, може бути зовсім непридатним для задач іншого типу.

Було запропоновано декілька функцій, які через свої властивості є тестовими для чисельних методів. Прикладом таких функцій є

- функція Розенброка:

$$Z = f(x_1, x_2) = 100(x_2 - x_1^*)^2 + (1 - x_1)^2, \hat{X} = (1, 1); \quad (9.1)$$

- функція Пауелла:

$$Z = f(x) = (x_1 + 10x_2)^2 + 5(x_3 - x_4)^2 + (x_1 - x_4)^4 + 10(x_1 - x_4)^4, \quad (9.2) \\ \hat{X} = (0, 0, 0, 0).$$

Будь-яка оптимізаційна процедура має ефективно розв'язувати задачі (9.1), (9.2) та інші тестові задачі.

Методи оптимізації поділяють на класи залежно від інформації, що використовується. Якщо на кожній ітерації використовується лише значення функцій, то метод називається методом нульового порядку. Якщо, крім того, буде потрібним обчислення перших похідних функцій, то мають місце методи першого порядку, при необхідності додаткового обчислення других похідних — методи другого порядку.

Розділ 9.2. Методи нульового порядку

Будемо вести мову про пошук мінімального значення функції n змінних.

У методах нульового порядку для визначення напрямку спуску не потрібне обчислення похідних цільової функції. Напрямок мінімізації повністю визначається послідовними обчисленнями значень функції. Зазначимо, що при розв'язанні задач безумовної оптимізації методи першого та другого порядку мають, як правило, більш високу швидкість збіжності, ніж методи нульового порядку. Але виникає проблема щодо обчислення перших та других похідних функцій багатьох змінних. В деяких випадках їх не можна отримати у вигляді аналітичних функцій. Визначення похідних за допомогою різних чисельних методів спричиняє помилки, які можуть обмежити застосування таких методів. Крім того, на практиці зустрічаються задачі, розв'язання яких можливе лише методами нульового порядку, наприклад, задачі мінімізації функцій із розривними першими похідними, Критерій оптимальності може задаватися не в явному вигляді, а системою рівнянь. У цьому випадку аналітичне або чисельне визначення похідних стає надто складним, а іноді й неможливим. Для розв'язання таких практичних задач оптимізації рекомендується застосувати методи нульового порядку.

Метод прямого пошуку (метод Хука-Дживса)

Пошук (наприклад мінімуму) складається з послідовності кроків дослідного пошуку навколо базисної точки, за яким у випадку успіху іде пошук за зразком.

Суть методу. Задається деяка початкова точка $X(0)$. Змінюючи компоненти вектора $X(0)$, обслідуються окіл обраної точки, в результаті чого знаходиться напрям, в якому зменшується функція $Z = f(X)$. У вибраному напрямі здійснюється спуск доти, поки значення функції зменшується. Після того, як у даному напрямі не знаходиться точка з меншим значенням функції, зменшується величина кроку спуску. Якщо послідовні дроблення кроку не призводять до зменшення функції, від вибраного напрямку спуску відмовляються і здійснюється нове обстеження околу і т. д.

Метод не потребує знання цільової функції в явному вигляді, дозволяє легко враховувати обмеження на окремі змінні, а також складні обмеження на допустиму область пошуку.

Недоліком методу є те, що у випадку сильно витягнутих, зігнутих ліній рівня цільової функції, він може не забезпечити рух до точки мінімуму.

Алгоритм методу прямого пошуку такий:

1. Задаються значення координат $x_j(0)$, $j = \overline{1, n}$ початкової точки $X(0)$, вектором зміщення координат ΔX у процесі обстеження околу, найменшим допустимим значенням ε компонент ΔX .
2. Покладаючи, що $X(0)$ є базисною точкою X^b , обчислюють $f(X^b)$.

3. Циклічно змінюють кожену координату x_j^b , $j = \overline{1, n}$, базисної точки X^b на величину $\Delta x_j, j = \overline{1, n}$, тобто $x_j(0) = x_j^b + \Delta x_j$, $x_j(0) = x_j^b - \Delta x_j$. При цьому обчислюються значення $f(X(k))$ і порівнюються з $f(X^b)$. Якщо $f(X(k)) < f(X^b)$, то відповідна координата $x_j, j = \overline{1, n}$ приймає нове значення, яке обчислюється за одним із наведених вище виразів. Такий крок вважають успішним. В протилежному випадку значення цієї координати залишається незмінним. Якщо після змінення останньої n -ї координати $f(X(k)) < f(X^b)$, то переходять до п.4. В протилежному випадку – до п.7.
4. Покладають, що $X(k)$ є новою базисною точкою X^b , обчислюють $f(X^6)$ і здійснюють пошук за зразком.
5. При пошуку за зразком використовується інформація. Отримана в процесі дослідження, і мінімізація функції завершується пошуком у напрямі, який задається зразком. Ця процедура здійснюється так: здійснюється спуск із точкою $X(k)$, $x_j(k+1) = 2x_j(k) - x_j^b, j = \overline{1, n}$, де x_j^b – координати попередньої базисної точки, обчислюється значення $f(X(k+1))$.
6. Як і в п.3, циклічно змінюють кожену координату точки $X(k+1)$, здійснюючи порівняння відповідних значень функції $f(X)$ із значенням $f(X(k+1))$, яке отримано в п.5. Після змінення останньої координати порівнюють відповідне значення функції $f(X(k+1))$ із значенням $f(X^6)$, отримано в п.4. Якщо $f(X(k+1)) < f(X^b)$, то переходять до п.4, в протилежному випадку – до п.7.
7. Порівнюють значення ΔX та ε . Якщо $\Delta X < \varepsilon$, то обчислення припиняють. Вважаючи $X(k+1)$ – розв'язком. У протилежному випадку зменшують значення ΔX на $\Delta X / 2$ і переходять до п.3.

На практиці задовільним є зменшення кроку (кроків) у десять разів від початкової довжини.

Метод деформованого многогранника (метод Нелдера–Міда)

Симплексний метод полягає в тому, що для мінімізації функції n -змінних $Z = f(X)$ у n -вимірному просторі будується многогранник, який має $n + 1$ вершину. Множина $(n + 1)$ -ї рівновіддаленої точки в n -вимірному просторі називається регулярним симплексом. Для випадку двох змінних регулярний симплекс буде рівнобічним трикутником, трьох змінних – тетраедром і т.д. Координати вершин початкового симплексу можна визначити за допомогою матриці R розміром $n \times (n + 1)$:

$$R = \begin{pmatrix} 0 & r_1 & r_2 & \cdots & r_2 \\ 0 & r_2 & r_1 & \cdots & r_2 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & r_2 & r_2 & \cdots & r_1 \end{pmatrix}, \quad (9.3)$$

де

$$r_1 = \left\lfloor \frac{l}{n\sqrt{2}} \right\rfloor (\sqrt{n+1} + n - 1),$$

$$r_2 = \left\lfloor \frac{l}{n\sqrt{2}} \right\rfloor (\sqrt{n+1} - 1)$$

– параметр, який ототожнюється з віддалю між двома вершинами. Елемент r_{ij} матриці R дорівнює i -й координаті j -ї вершини.

Пошук мінімуму функції симплексним методом ведеться таким чином. Установлюються координати вершин симплексу і в кожній з них обчислюється значення функції $f(X)$. Визначається вершина з найбільшим значенням $f(X)$. Через цю вершину й центральну точку симплексу проводиться пряма, на якій з деяким віддаленням від центру установлюється нова вершина. Потім вершина з найбільшим значенням $f(X)$ вилучається, а новий симплекс будується на вершинах, що залишилися, та новій вершині. Цей метод, використовує регулярні симплекси, може зациклюватися, має певні труднощі в ярих ситуаціях.

Метод Нелдера-Міда усуває ці недоліки. Координати вершин початкового симплексу можуть бути призначені або обчислені за (9.3). На наступних етапах конфігурація симплексу змінюється. Симплекс втрачає регулярність. Тому цей метод має другу назву – метод деформованого многогранника.

Введемо такі позначення:

$$X(j, k) = (x_1(j, k), \dots, x_i(j, k), \dots, x_n(j, k)),$$

де $j = 1, n+1$, $k = 0, 1$ – j -та вершина многогранника на k -му етапі пошуку;

$X(h, k)$ – вершина, в якій значення функції максимальне, тобто:

$$f(X(h, k)) = \max \{f(X(1, k)), \dots, f(X(n+1, k))\},$$

$X(l, k)$ – вершина, в якій значення функції мінімальне, тобто:

$$f(X(l, k)) = \min \{f(X(1, k)), \dots, f(X(n+1, k))\};$$

$X(n+1, k)$ – «центр ваги» всіх вершин за виключенням $X(h, k)$.

Координати «центру ваги» обчислюються за формулою

$$x_i(n+2, k) = \frac{1}{n} \left(\sum_{j=1}^{n+1} x_j(j, k) - x_i(h, k) \right), \quad i = \overline{1, n},$$

де i – номер координати.

У методі Нелдера-Міда над многогранником виконуються такі операції (рис.9.1):

1. **Відбиття**, тобто проектування $X[h, k]$ через центр ваги відповідно із співвідношенням:

$$X(n+3, k) = X(n+2, k) + \alpha(X(n+2, k) - X(n, k)),$$

де $\alpha > 0$ - коефіцієнт відбиття.

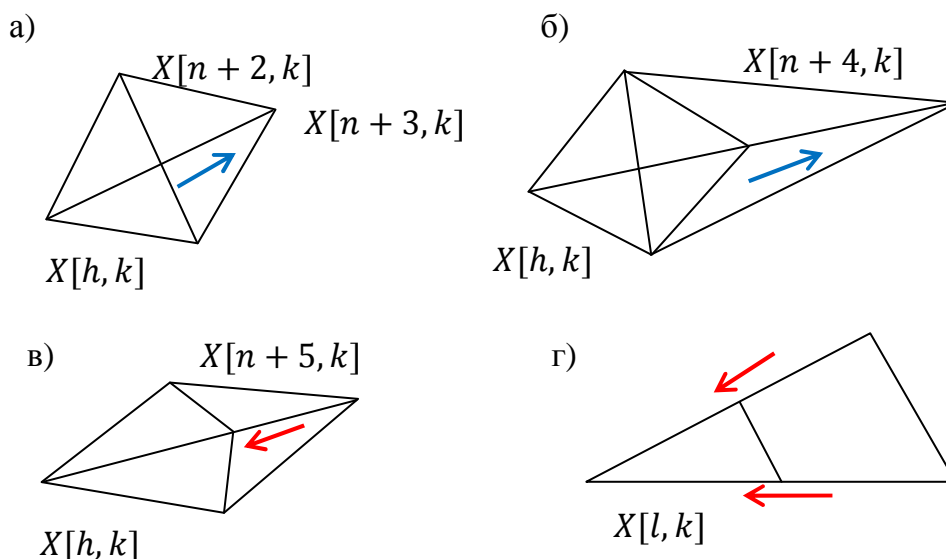
2. **Розтяг** застосовується у тому випадку, коли відбиття є вдалим, тобто коли

$$f(X(n+3, k)) \leq f(X(l, k)).$$

При цьому вектор $X(n+3, k) - X(n+2, k)$ розтягується і координати нової вершини визначаються як:

$$X[n+4, k] = X[n+2, k] + \gamma(X[n+3, k] - X[n+2, k]),$$

де $\gamma > 1$ - коефіцієнт розтягу.



**Рис. 9.1 Операції над многогранником:
а) відбиття; б) розтяг; в) стиск; г) редукція**

3. **Стиск** виконується, якщо в результаті відбиття значення функції в точці $X[n+3, k]$, більше від значень у всіх інших вершинах симплексу за винятком вершини $X[h, k]$, тобто:

$$f(X[n+3, k]) > f(X[j, k])$$

для всіх $j \neq h$.

Тоді вектор $X[h, k] - X[n+2, k]$ стискується так, що

$$X[n+5, k] = X[n+2, k] + \beta(X[h, k] - X[n+2, k])$$

де $0 < \beta < 1$ - коефіцієнт стиску.

4. **Редукція**, тобто стиск симплексу у два рази відносно до вершини $X[l, k]$. Редукція застосовується у випадку, коли:

$$f(X[n+3, k]) > f(X[h, k])$$

виконується за формулою:

$$X[j, k] = X[l, k] + 0.5(X[j, k] - X[l, k]), \quad j = \overline{1, n+1}.$$

Зауваження 9.1.

Як правило, значення коефіцієнтів відбиття, розтягу та сти-
ску треба приймати:

$$\alpha = 1; \quad 0.4 \leq \beta \leq 0.62; \quad 8 \leq \gamma \leq 3.0,$$

хоча Недлер і Мід рекомендують брати:

$$\alpha = 1; \quad \beta = 0.5; \quad \gamma = 2.$$

Умова припинення пошуку записується у вигляді:

$$\sigma = \left\{ \frac{1}{n+1} \sum_{j=1}^{n+1} \left(f(X[j, k]) - f(X[n+2, k]) \right)^2 \right\}^{0.5} \leq \varepsilon \quad (9.4)$$

де $\varepsilon > 0$ – мале число, яке визначає ε окіл пошуку екстремуму.

Алгоритм методу деформованого многогранника об'єднує такі кроки:

Крок 1. Установлення координат вершин початкового симплексу і обчислення в кожній із них функції $f(X)$.

Крок 2. Визначення вершин з найбільшим і найменшим значеннями $f(X)$.

А також координат центру ваги симплексу: $X[h, k]$, $X[l, k]$, $x_i[n+2, k]$, $i = \overline{1, n}$.

Крок 3. Відбиття: обчислення координат відбитої вершини $x_i[n+2, k]$, $i = \overline{1, n}$ і значення функції $f(X[n+3, k])$.

Крок 4. Якщо $f(X[n+3, k]) > f(X[l, k])$, то обчислення продовжується з кроку 7.

Крок 5. Розтяг: обчислення координат $x_i[n+4, k]$, $i = \overline{1, n}$ і значення $f(X[n+4, k])$.

Крок 6. Приймається

$$X[h, k] = \begin{cases} X[n+4, k], & \text{якщо } f(X[n+4, k]) < f(X[l, k]) \\ X[n+3, k], & \text{якщо } f(X[n+4, k]) \geq f(X[l, k]) \end{cases}$$

і здійснюється перехід до кроку 12.

Крок 7. Перевірка умов $f(X[n+3, k]) > f(X[j, k])$ для всіх $j \neq k$. Якщо ця нерівність не виконується, то приймається $X[h, k] = X[n+3, k]$ і пошук продовжується з кроку 12.

Крок 8. Приймається

$$X[h, k] = \begin{cases} X[n+3, k], & \text{якщо } f(X[n+3, k]) \geq f(X[h, k]) \\ X[h, k], & \text{якщо } f(X[n+3, k]) < f(X[h, k]) \end{cases}.$$

Крок 9. Стиск: обчислення координат

$$x_i[n+5, k], \quad i = \overline{1, n}, \quad f(X[n+5, k]).$$

Крок 10. Якщо $f(X[n + 5, k]) \leq f(X[h, k])$, то приймається $X[h, k] = X[n + 5, k]$ і пошук проводиться з кроку 12.

Крок 11. Редукція.

Крок 12. Якщо умова припинення пошуку (9.4) не виконується, то індекс k збільшується на одиницю і пошук мінімуму повторюється з кроку 2. Якщо умова (9.4) виконується, то усі значення функцій є дуже близькими одне до одного, тому вони, можливо, лежать поблизу точки мінімуму функцій $X[l, k]$. Виходячи з цього, такий критерій збіжності є розумним.

Розділ 9.3. Прямі та непрямі методи умовної оптимізації

В загальному випадку методи поділяються на прямі та непрямі. Прямі методи оперують безпосередньо з вихідними задачами оптимізації і генерують послідовність точок $\{X[k]\}$ таких, що $f(X[k + 1]) < f(X[k])$. В силу цього такі методи часто називають методами спуску. Математично перехід на певному k -кроці ($k = 0, 1, 2, \dots$) від точки $X[k]$ до точки $X[k + 1]$ можна записати так:

$$X[k + 1] = X[k] + \alpha_k P[k], \quad (9.5)$$

де $P[k]$ – вектор, який визначає напрям спуску; α_k – довжина кроку вздовж даного напрямку.

При цьому в одних алгоритмах прямих методів точки $X[k]$ вибираються так, щоб для них виконувалися всі обмеження задачі, в інших ці обмеження можуть порушуватися на деяких або всіх ітераціях.

Таким чином, в прямих методах при виборі напрямку спуску обмеження, що визначають допустиму область G , враховують в явному вигляді.

Непрямі методи зводять вихідну задачу нелінійного програмування до послідовності задач безумовної оптимізації певних допоміжних функцій. При цьому обмеження вихідної задачі враховуються в явному вигляді.

Розглянемо деякі алгоритми прямих методів.

Прямі методи умовної оптимізації

1. Метод проекції градієнта

Нехай маємо таку загальну задачу: мінімізувати функцію $Z = f(X)$ за наявності m обмежень – нерівностей $g_i(X) \leq 0$, $i = 1, m$ (обмеження $g_i(X) \leq b_i$ можна записати так $g_i(X) - b_i \leq 0$). За початкову вибирається деяка точка допустимої області G . Якщо $X[0]$ – внутрішня точка множини G , то розглядуваний метод є звичайним градієнтним методом:

$$X[k + 1] = X[k] - \alpha_k \nabla f(X[k]), \quad k = 0, 1, 2, \dots \quad (9.6)$$

де $\nabla f(X[k])$ – градієнт функції $f(X)$ в точці $X[k]$.

Після виходу на межу області G в якійсь точці $X[k]$ рух в напрямі антиградієнта – $\nabla f(X[k])$ може вивести за межі допустимої множини (рис. 9.2). Тому антиградієнт проектується на лінійний многовид M , який апроксимує ділян-

ку межі в околі точки $X[k]$. Рухаючись в напрямі проекції вектора $-\nabla f(X[k])$ на многовид M , відшукується нова точка $X[k + 1]$, в якій $f(X[k + 1]) < f(X[k])$. $X[k + 1]$ приймається за вихідне наближення і процес продовжується.

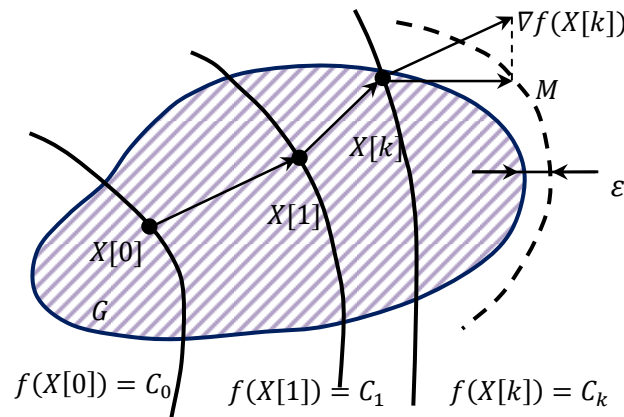


Рис. 9.2. Геометрична інтерпретація методу проекції градієнта

Розглянемо більш детально цю процедуру. В точці $X[k]$ частина обмежень – нерівностей задовольняються як рівності

$$g_i(X) = 0, \quad i = \overline{1, l}, \quad l < m.$$

Такі обмеження називають активними. Позначимо через I множину індексів i ($1 \leq i \leq l$) цих обмежень. Їх рівняння відповідають гіперповерхням, що утворюють межу області G в околі точки $X[k]$. В загальному випадку ця межа буде нелінійною. Обмеження $g_i(X)$, $i \in I$, апроксимуються гіперповерхням, які дотикаються до них у точці $X[k]$:

$$\sum_{j=1}^n \frac{\partial g_i(X[k])}{\partial x_j} (x_j - x_j[k]) = 0, \quad i \in I \quad (9.7)$$

Отримані гіперповерхні обмежують деякий многогранник M , який апроксимує допустиму область G в околі точки $X[k]$. Проекція $p[k]$ антиградієнта $-\nabla f(X[k])$ на многогранник M обчислюється за формулою:

$$P[k] = P[-\nabla f(X[k])] = 0, \quad i \in I \quad (9.8)$$

де P – оператор ортогонального проектування, який визначається виразом:

$$P = E - A^T (AA^T)^{-1} A,$$

де E – одинична матриця розміру $n \times n$,

A – матриця розміру $l \times n$ – якобіан функції $g_i(X)$, який обчислений у точці $X[k]$

$$A = [\alpha_1, \dots, \alpha_i, \dots, \alpha_l]^T,$$

$$\text{де } \alpha_i = \left(\frac{\partial g_i(X[k])}{\partial x_1}, \dots, \frac{\partial g_i(X[k])}{\partial x_n} \right), \quad i = \overline{1, l}.$$

Далі здійснюється спуск у обраному напрямі.

$$X[k+1] = X[k] + \alpha_k p[k].$$

Можна показати, що точка $X[k+1]$, буде розв'язком задачі мінімізації функції $f(X)$ в області G тоді і тільки тоді, коли $P[-\nabla f(X[k])] = 0$, тобто

$$-\nabla f(X[k]) = \sum_{i=1}^l u_i \alpha_i \text{ та } U = (A^T A)^{-1} A^T (-\nabla f(X[k])) \geq 0.$$

Ця умова означає, що антиградієнт $-\nabla f(X[k])$ цільової функції є лінійною комбінацією з невід'ємними коефіцієнтами градієнтів обмежень $g_i(X) = 0$.

Відповідно до викладеного вище алгоритм методу проекції градієнта складається з таких дій:

1. У точці $X[k]$ визначається напрям спуску $p[k]$.

Оскільки точка $X[k] \in G$ і відомі активні обмеження $g_i(X) = 0, i \in I$, то обчислюються $-\nabla f(X[k])$ і визначається проекція $P[-\nabla f(X[k])]$. При цьому можливі два випадки:

а) $P[-\nabla f(X[k])] \neq 0$, тоді за напрям спуску приймається отримана проекція,

б) $P[-\nabla f(X[k])] = 0$, тобто

$$\nabla f(X[k]) = \sum_{i=1}^l u_i \alpha_i.$$

Якщо усі $u_i \geq 0, i \in I$, то точка $X[k]$ буде розв'язком задачі. Якщо ж деяка з компонент $u_q < 0$, то q -й градієнт вилучається з матриці A і породжується нова матриця проектування P , яка визначає новий напрям спуску.

1. Для визначення величини кроку α_k цільова функція мінімізується у напрямі $p[k]$ за умов виконання обмежень задачі із заданою точністю, що задається певним додатнім числом ε . Вважається, що точка X задовольняє умовам задачі із заданою точністю, якщо $g_i(X) = \varepsilon, i = \overline{1, m}$. Величина кроку α_k визначається розв'язком задачі вигляду:

$$\begin{aligned} f(X[k] + \alpha p[k]) &\rightarrow \min, \\ g_i(X[k] + \alpha p[k]) &\leq \varepsilon, \quad i = \overline{1, m}. \end{aligned}$$

2. Наступна точка обчислюється за формулою

$$X[k+1] = X[k] + \alpha p[k].$$

Ознакою збіжності є наближення до нуля векторів $p[k]$.

Розглянутий метод є аналогом градієнтних методів для розв'язання задач на безумовний екстремум, тому він також поволі збіжний.

2. Комплексний метод Бокса

Цей метод є модифікацією методу деформованого многогранника, що дозволяє врахувати обмеження нерівності. Задачі, що розв'язується, полягає в мінімізації функції

$$Z = f(X),$$

де $X = (x_1, \dots, x_n)$ визначається явними обмеженнями

$$\alpha_j \leq x_j \leq b_j, \quad j = \overline{1, n},$$

а також неявними обмеженнями

$$g_i(X) \leq 0, \quad i = \overline{1, m}.$$

Якщо цільова функція задачі опукла і функції $g_i(X)$ також опуклі, то задача має єдиний розв'язок. Значення α_j та b_j є нижньою та верхньою межами змінних.

Вибираємо q точок, які задовольняють обмеженням задачі, а також обчислимо в цих точках цільову функцію. Побудуємо многогранник, який має $q > n + 1$ вершин (Бокс знайшов, що $q = 2n$). Цей многогранник називається комплексом.

Введемо такі позначення:

$$X[i, k] = (x_1[i, k], \dots, x_j[i, k], \dots, x_n[i, k]),$$

де $i = \overline{1, q}$, $k = 0, 1, 2, \dots$ – i -а вершина комплексу на k -му кроці,

$X[h, k]$ – вершина, в якій значення цільової функції максимальне.

Тобто

$$f(X[h, k]) = \max \{X[1, k], \dots, f(X[q, k])\},$$

$X[k, k]$ – центр ваги усіх вершин за виключенням $X[h, k]$.

Координати центру ваги обчислюються за формулою

$$x_j[l, k] = \frac{1}{q} \left(\sum_{i=1}^q x_j[i, k] - x_j[h, k] \right), \quad j = \overline{1, n}. \quad (9.9)$$

За першу вершину початкового комплексу вибирається деяка допустима точка $X[1, 0]$. Координати інших $q - 1$ вершин комплексу визначаються співвідношенням

$$x_j[i, 0] = \alpha_j + r_j (b_j - \alpha_j) \quad (9.10)$$

для $j = \overline{1, n}$, $i = 2, q, r_j$ - псевдовипадкові числа, які рівномірно розподілені в проміжку $[0, 1]$. Точки, вибрані з цим для даного j , будуть автоматично задовольняти обмеження $\alpha_j \leq x_j \leq b_j$, але обмеження $g_i(X) \leq 0$ можуть бути порушені. В цьому випадку недопустима точка замінюється новою, такою, що лежить в середині відрізка. Який сполучає точку з центром ваги вибраних допустимих вершин, тобто формується точка

$$X[i, k] = \frac{X[i, k] + X[l, k]}{2} \quad (9.11)$$

Ця процедура може повторюватися доти, поки точка не стане допустимою. Якщо функції $g_i(X)$ опуклі, то, врешті-решт, обмеження виконуватимуться і комплекс утворюватиметься допустимими точками.

Розглянемо ітераційну процедуру комплексного методу, в якій здійснюється відшукання мінімуму переміщення за напрямом до мінімуму всередині області обмежень. Тут треба виконати такі кроки:

1. Знаходимо точку з найбільшим значенням функції $X[h, k]$ та центр ваги $(q - 1)$ точок, що залишилися.
2. Як і в методі деформованого многогранника, замінюємо вершину $X[h, k]$ шляхом відображення відносно центру ваги $X[l, k]$, обчислюється за формулою

$$X[p, k] = (\alpha + 1)X[l, k] + \alpha X[h, k],$$

де $\alpha > 0$ – деяка константа, яку називають коефіцієнтом відображення. Найбільш задовільні результати значення $\alpha = 1.3$.

3. Перевіряємо, чи буде отримана точка $X[p, k]$ допустимою.

а) якщо $X[p, k]$ не буде допустимою та не виконується обмеження для α_j , то покладемо $x_j[p, k] = \alpha_j + 10^{-6}$, якщо ж не виконується обмеження для b_j , то покладемо

$$x_j[p, k] = b_j - 10^{-6};$$

б) якщо обмеження не виконуються, то точку $X[p, k]$ пересувають на половину відстані між $X[p, k]$ та $X[l, k]$, тобто

$$X[p, k](\text{нове}) = \frac{X[p, k] + X[l, k]}{2}.$$

Ця процедура може повторюватися декілька разів, поки не буде отримана допустима точка.

4. Якщо точка $X[p, k]$ допустима, то обчислюється значення функції в цій точці $f(X[p, k])$ і порівнюється з найбільшим значення функції $f(X[h, k])$. Якщо $f(X[p, k]) > f(X[h, k])$, то точка $X[p, k]$ зсувається до центру $X[l, k]$ на половину відстані між ними, тобто

$$X[p, k]_{\text{(нове)}} = \frac{X[p, k] + X[l, k]}{2}$$

і процес повертається до кроку 3.

5. Якщо $f(X[p, k]) < f(X[h, k])$, то точка $X[h, k]$ замінюється точкою $X[p, k]$ і точки значення функцій комплексу знову упорядковуються.

6. Критерій завершення: обчислення закінчуються, коли значення цільової функції мало змінюється протягом п'яти послідовних ітерацій, тобто

$$\left| f(X[l, k+1]) - f(X[l, k]) \right| \leq \varepsilon, \quad k = 1, 2, \dots, 5,$$

де ε – задана константа.

В цьому випадку середнє арифметичне п'яти останніх ітерацій вважають розв'язком задачі нелінійного програмування. В протилежному випадку треба повернутися до кроку 1 та повторити процедуру.

Комплексний метод можна застосовувати до широкого кола задач з обмеженнями. Він простий та надійний у роботі. На кожному кроці метод використовує тільки інформацію про значення цільової функції та функцій обмежень задачі. Якщо цільова функція опукла і, крім того, опукла область обмежень, то застосування методу буде успішним, хоча певні особливості задачі можуть вимагати певної модифікації критерію завершення.

Непрямі методи умовної оптимізації (методи штрафних функцій)

Методи штрафних функцій відносяться до непрямих методів розв'язування ЗНП. Методи полягають у зведенні задачі математичного програмування до однієї чи кількох задач безумовної оптимізації деяких спеціальним способом побудованих допоміжних функцій. У загальному вигляді допоміжну функцію можна записати так:

$$F(X, \alpha) = f(X) + \Phi(X, \alpha),$$

де $f(X)$ – мінімізуюча цільова функція задачі,

$\Phi(X, \alpha)$ – "штрафна" функція, параметр $\alpha > 0$.

Точку безумовного мінімуму функції $F(X, \alpha)$ позначимо через $X(\alpha)$.

Залежно від вигляду $\Phi(X, \alpha)$ розрізняють методи внутрішніх або зовнішніх штрафних функцій, які інколи називають «бар'єрними» функціями.

1. Методи внутрішніх штрафних функцій

Ці методи застосовуються при розв'язуванні ЗНП з обмеженнями – нерівностями типу $g_i(X) \leq 0, i = 1, m$. Функції $\Phi(X, \alpha)$ підбирають так, щоб їх значення необмежено зростало при наближенні до межі допустимої області G (рис. 9.3.).

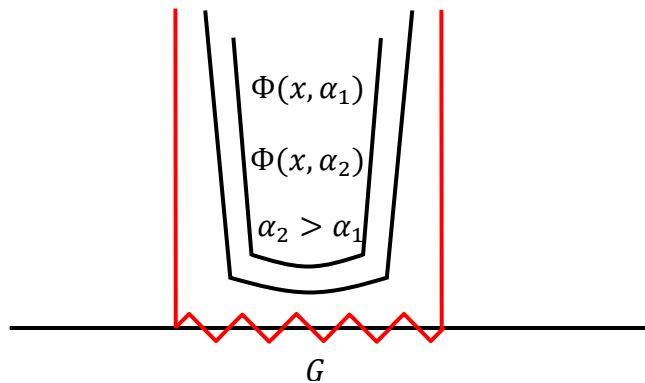


Рис. 9.3. Геометрична інтерпретація методу внутрішніх штрафних функцій

Наближення до межі "штрафується" різким збільшенням значення функції $F(X, \alpha)$. На межі G побудовано "бар'єр", який перешкоджає порушенню обмежень у процесі безумовної оптимізації $F(X, \alpha)$. Пошук мінімуму допоміжної функції треба починати із внутрішньої точки області G . Тоді в процесі оптимізації траєкторія спуску ніколи не вийде за межі допустимої області. Усі перелічені особливості функції $\Phi(X, \alpha)$ визначили найменування розглядуваної групи методів.

Отже, внутрішня штрафна функція $\Phi(X, \alpha)$ може бути визначена так:

$$\Phi(X, \alpha) = \begin{cases} \infty, & \text{якщо } X \notin G, \\ 0, & \text{якщо } X \in G, X \notin \partial G, \alpha \rightarrow 0, \\ \infty, & \text{якщо } X \in G, X \rightarrow \partial G. \end{cases} \quad (9.12)$$

де ∂G - межа області G .

Загальний вигляд внутрішньої штрафної функції

$$\Phi(X, \alpha) = \alpha \sum_{i=1}^m \varphi_i(g_i(X)), \quad (9.13)$$

де φ_i - неперервні диференційовані функції. Які визначаються обмеженнями-нерівностями вихідної ЗНП.

Допоміжна функція $F(X, \alpha)$ при цьому має вигляд:

$$F(X, \alpha) = f(X) + \alpha \sum_{i=1}^m \varphi_i(g_i(X)). \quad (9.14)$$

Вона визначена в області G і необмежено зростає, коли $g_i(X) \rightarrow 0$ для деякого i . Прикладами внутрішніх штрафних функцій для задач мінімізації є такі:

$$\Phi(X, \alpha) = \alpha \sum_{i=1}^m \frac{1}{g_i(X)}; \quad \Phi(X, \alpha) = -\alpha \sum_{i=1}^m \ln[g_i(X)].$$

Алгоритм методу такий. За початкову точку вибираємо довільну внутрішню точку області G . Задаємо деяку монотонно спадну збіжну до нуля послідовність $\{a_k\}, k = 1, 2, \dots$ додатних чисел. Для першого елемента α_i цієї послідов-

ності розв'язуємо задачу безумовної мінімізації функції $F(X, \alpha_1)$, у результаті чого визначаємо точку $X(\alpha_1)$. Ця точка використовується за початкову для розв'язування задачі пошуку мінімуму функції $F(X, \alpha_2)$, де $\alpha_2 \leq \alpha_1$ і т.д. Отже, розв'язуємо послідовність задач безумовної мінімізації функції $F(X, \alpha_k), k = 1, 2, \dots$, причому розв'язок попередньої задачі $X(\alpha_k)$ використовуємо за початкову точку для пошуку наступного вектора $X(\alpha_{k+1})$. Послідовність точок $X(\alpha_k)$ збігається до оптимального розв'язку вихідної задачі – локального мінімуму X^* .

Обчислення припиняються, коли

$$\begin{aligned} |f(X[k]) - f(X[k+1])| &\leq \varepsilon, \\ \|X[k] - X[k+1]\| &\leq \beta, \end{aligned}$$

де ε, β – задані числа, які визначають точність обчислень.

Розглянутий метод внутрішніх штрафних функцій має такі властивості

- 1) $\lim_{k \rightarrow \infty} \alpha_k \sum_{i=1}^m \varphi_i(g_i(X[k])) = 0$;
- 2) $\lim_{k \rightarrow \infty} f(X[k]) = f(X^*)$, $f(X[k])$ – монотонно спадає;
- 3) $\lim_{k \rightarrow \infty} F(X[k], \alpha_k) = f(X^*)$.

Ці властивості є вірними для задач, які мають неперервні функції і локальні мінімуми всередині області G .

2. Методи зовнішніх штрафних функцій

Функції $\Phi(X, \alpha)$ вибирають так, що їх значення дорівнює нулю всередині і на межі допустимої області G , а поза нею – додатні і зростають тим більше, чим сильніше порушуються обмеження (рис. 9.4.). Отже, тут «штрафується» віддалення від допустимої області. Зовнішня графічна функція $\Phi(X, \alpha)$ в загальному вигляді можна визначити так:

$$\Phi(X, \alpha) = \begin{cases} 0, & \text{якщо } X \in G, \\ \infty, & \text{якщо } X \notin G, \alpha \rightarrow \infty. \end{cases}$$

Пошук мінімуму допоміжної функції $F(X, \alpha)$ починається із довільної точки.

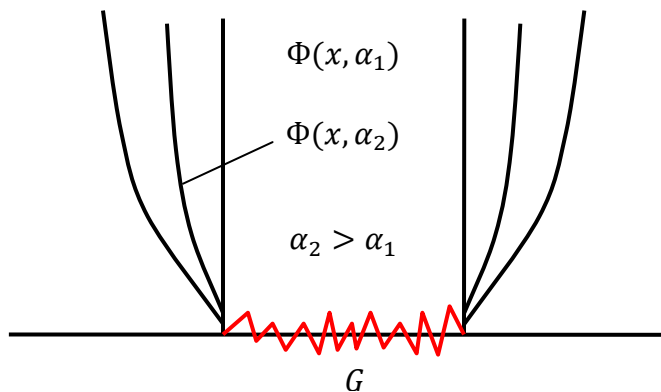


Рис. 9.4. Геометрична інтерпретація методу зовнішніх штрафних функцій

У більшості випадків вона є недопустимою, тому траєкторія спуску розміщується частково за допустимою областю. Якщо мінімум цільової функції знаходиться на межі допустимої області. То ця траєкторія повністю лежить за областю G . Перелічені особливості функції $\Phi(X, \alpha)$ визначили назву даної групи методів. Загальний вигляд зовнішньої функції

$$\Phi(X, \alpha) = \alpha \left(\sum_{i=1}^m \varphi_i(g_i(X)) \right) \quad (9.15)$$

де φ_i – функція, яка визначається обмеженнями-нерівностями ЗНП

Допоміжна функція $F(X, \alpha)$ при цьому така:

$$F(X, \alpha) = f(X) + \alpha \left(\sum_{i=1}^m \varphi_i(g_i(X)) \right). \quad (9.16)$$

Прикладами зовнішніх штрафних функцій можуть бути такі:

$$\Phi(X, \alpha) = \left\{ \max[0, g_i(X)] \right\}^2, \quad \Phi(X, \alpha) = \alpha \exp \left\{ \frac{1}{\alpha} \sum_{i=1}^m \max[0, g_i(X)] \right\}.$$

Тут

$$\max[0, g_i(X)] = \begin{cases} 0, & \text{якщо } g_i(X) \leq 0, \\ g_i(X), & \text{якщо } g_i(X) > 0. \end{cases}$$

Алгоритм методу зовнішніх штрафних функцій формулюється також, як і алгоритм внутрішніх штрафних функцій і має аналогічні властивості. Однак в цьому випадку не потрібно, щоб точка $x[0] \in G$, а послідовність $\{\alpha_k\}, k = 1, 2, \dots$, додатних чисел була монотонно зростаючою.

Аналіз методів штрафних функцій дає такі висновки. Методи внутрішніх функцій ведуть пошук розв'язку, не виходячи за межі допустимої області. Це важливо, коли цільова функція чи обмеження не визначені за границями допустимої множини. Крім того, припиняючи обчислення, ми завжди отримуємо допустимий розв'язок. Але для визначення будь-якої початкової точки іноді треба розв'язати задачу, яка порівняна за складністю із вихідною задачею. Методи зовнішніх штрафних функцій забезпечують розв'язування із довільної точки. Загальним недоліком методів є складність допоміжної функції, яка може мати

яристу структуру. Степінь яристості збільшується із збільшенням α . При великих значеннях α точність обчислень мінімуму $F(X, \alpha)$ значно зменшується через помилки округлення.